

Информационно-вычислительная среда Mathcell для моделирования живой клетки

©2007 В. Лахно*, Н. Назипова, В. Ким, С. Филиппов, Н. Фиалко, Д. Устинин, А. Теплухин, Г. Тюльбашева, А. Зайцев, М. Устинин
Институт математических проблем биологии РАН, Пуцзино, Моск. обл., Россия

Аннотация. Описывается созданный коллективом информационно-вычислительный портал «Математическая клетка» (<http://www.mathcell.ru>), включающий в себя 3D-модель эукариотической клетки, снабженную возможностями навигации во внутреннем пространстве и связанную с информационной составляющей портала, энциклопедию знаний по математической биологии и коллекцию оригинальных моделей отдельных клеточных процессов.

Проблема разработки адекватной модели живой клетки давно волнует умы ученых. Долгое время моделировались отдельные клеточные процессы (кинетиические превращения, процессы роста клеток и их колоний и др.), в СССР первые работы, делавшие робкие попытки построить математическую модель клетки целиком появились в середине 80-х годов XX века [1]. Однако большого распространения эти работы не получили. Лишь достижения последних годов в области биологии и компьютерных и коммуникационных технологий дали новый импульс активизации этих работ. Впервые появилась возможность приступить к решению задачи построения интегрированной компьютерной модели живой клетки, клетки *in silico* [2, 3]. В этой области идет активная совместная работа биологов и математиков, положение меняется очень быстро. В настоящее время моделирование клеточных процессов находится на самом переднем крае клеточной и математической биологии.

Самым первым проектом такого рода был проект E-CELL [4], появление и развитие которого связывают с личностью директора Institute for Advanced Biosciences at Keio University. М. Tomita первым в 1996 году решил создать единую модель самовоспроизводящейся живой клетки [5]. Первым объектом для такого моделирования была выбрана бактерия *Mycoplasma genitalium*, геном которой, как известно, минимален и состоит из 127 генов. Первый опыт показал, что для успешного решения этой задачи необходимо большое количество экспериментальных данных, которых к тому времени было недостаточно. Проект развивается в направлении получения новых данных и их интеграции в модели функционирования человеческих эритроцитов, митохондрий, хемотаксиса у *E. Coli* и др.

Другим известным проектом по моделированию клетки является Virtual Cell (VCell), программный комплекс, созданный The National Resource for Cell Analysis and Modeling (NRCAM) [6]. Ресурс представляет собой программную платформу для моделирования клеточных процессов, она связывает биохимические и электрофизиологические данные, описывающие отдельные реакции, с экспериментальными микроскопическими данными, описывающими их внутриклеточное расположение. Клеточные физиологические события могут описываться эмпирическими геометрическими соображениями. Virtual Cell состоит из двух компонент – биологической и математической. Исследователь может создавать

* lak@impb.psn.ru

биологические объекты, для которых программа создает математический код, необходимый для имитации деятельности этих объектов. С помощью программного обеспечения Virtual Cell можно создавать действующие модели отдельных клеточных процессов.

Существует ряд других программных комплексов для моделирования [7-13], некоторые из них пользуются ресурсами вычислительной грид-инфраструктуры [14, 15]. Для того, чтобы разные компьютерные программы, созданные для различных операционных систем и для различных вычислительных платформ могли обмениваться между собой данными были созданы специальные стандарты для записи моделей.

Первым таким стандартом для представления в машинно-читаемом формате сетей биохимических реакций был SBML (Systems Biology Markup Language) [16], он предназначался для записи метаболических схем, путей передачи сигналов в клетках, регуляторных сетей.

Стандарт представления биологических данных SBML основан на стандарте XML. Магистральное направление развития Web-технологий в последнее время связано с новым расширяемым языком разметки документов XML (Extensible Markup Language) [17]. Стандарт этого языка был принят консорциумом W3C [18] в феврале 1998 года и в последнее время получает широкое распространение. Этот стандарт позволяет отделить данные от способа их представления. Это делает возможным структурирование данных, программирование операций над данными и их автоматическое редактирование; что, в свою очередь, позволяет передавать информацию самыми различными способами любым цифровым устройствам, а также организовать взаимодействие приложений и предоставление интерактивных, взаимодействующих друг с другом веб-сервисов.

В отличие от своего предшественника - языка HTML, XML является "открытым" языком. В нем не фиксируется множество тегов, используемых для разметки документов. Создатель XML-документа или набора документов может сам конструировать необходимые ему теги, определяя их с помощью метасредств языка XML - набора синтаксических конструкций, называемых декларациями разметки (Markup Declaration).

Язык XML позволяет описывать логическую структуру документов в терминах составляющих их элементов. Такие описания - Document Type Definition (DTD) и XML Schema - могут поддерживаться в глобальной среде независимо от экземпляров XML-документов и могут унифицироваться в интересах различных сообществ пользователей. Благодаря этому можно создавать новые языки разметки, адекватные потребностям различных сфер применения.

Консорциумом W3C разработаны различные специализированные стандарты, например язык разметки математических текстов MathML (Mathematical Markup Language). MathML, описывая совокупность XML-функций, позволяет осуществлять в WWW многочисленные математические операции. MathML может использоваться с целью кодирования и представления математического содержания для последующей высококачественной визуальной интерпретации для приложений, в которых основную роль играет семантика.

С 2000-го года существует международное сообщество SBML Team, оно поддерживается правительственными программами США, Японии и Англии. Сообщество создано для разработки и совершенствования стандарта представления моделей и программного обеспечения для работы с этим форматом. Стандарт SBML стал активно использоваться разработчиками программного обеспечения для исследования метаболических аспектов клетки [19-21], в настоящее время существует более 120 программных комплексов, поддерживающих стандарт SBML. Среди них много сервисных программ, например, для визуализации метаболических сетей, для графического задания новых сетей передачи сигнала и экспорта в SBML и др. Все они

перечислены на сайте сообщества SBML [16]. Создатели стандарта записали ряд известных к тому времени в литературе метаболических сетей в библиотеку моделей SBML. А по мере роста популярности этого стандарта и под влиянием появившегося позднее стандарта CellML, который обладал большими возможностями и позволял записывать математические формулы, разработчики стандарта дополнили его программным пакетом MathSBML [22] для работы с SBML-моделями. Он позволяет смотреть модели на экране (в формате SBML), конвертировать их в системы обыкновенных дифференциальных уравнений для расчета и визуализации результатов в среде Mathematica, переводить модели в другие форматы представления и является программным комплексом с открытым кодом.

Другой формат представления моделей клеточных процессов, имеющий в своей основе технологию XML и некоторое время конкурировавший с SBML, основан в Институте биоинженерии при Университете Окленда, Новая Зеландия. Специализированный стандарт обмена данными в области математического моделирования CellML [23-25] использует специализированные надстройки MathML, RDF, Dublin Core Metadata Set. Формат CellML получился довольно универсальным, он позволяет представлять математические формулы, графические схемы и диаграммы, данные о публикациях. Авторы стандарта создали по данным литературных источников свою коллекцию моделей, начиналась эта коллекция с электрофизиологических моделей и моделей передачи (трансдукции) сигналов, позже стала включать в себя модели метаболизма, иммунологические и др. модели.

Надо сказать, что, формат CellML не так сильно распространен, как SBML, но он сыграл свою положительную роль в совершенствовании формата представления моделей и включении в него необходимых компонент, которые сделали SBML соответствующим современным требованиям. Наличие большого количества разнофункционального программного обеспечения для работы с моделями, представленными в формате SBML, сделало этот стандарт записи биологических схем и взаимосвязей между их компонентами главенствующим в мире. Это привело к тому, что многие обособленные форматы представления данных, в первую очередь метаболических, в настоящее время конвертируются в формат SBML [26] и собираются в единую базу данных, которая принадлежит EMBL.

База данных BioModels [27, 28], которая аккумулировала все крупнейшие в мире разрозненные коллекции моделей, в числе прочих коллекции SBML и CellML, поддерживает наиболее известные на сегодняшний день форматы представления данных, в том числе и не упомянутые здесь [см., например, 29]. Возможно, когда-нибудь из разрозненных моделей различных клеточных процессов, представленных в этой базе данных, можно будет собрать общую единую модель клетки.

Мы представляем российский проект «Математическая клетка» (<http://www.mathcell.ru>), который идейно отличается от известных на сегодняшний день ресурсов.

Живые организмы представляют собой либо отдельные клетки, либо многоклеточные образования. Рассматривая клетку в электронный микроскоп, можно увидеть в ней длинные цепи сложной формы из сотен атомов – белки и нуклеиновые кислоты (ДНК и РНК). В клетке имеется множество типов белков, среди них ферменты, действующие как катализаторы различных метаболических процессов, и структурные белки, являющиеся ее строительным материалом. В клетках высших организмов имеется множество других белков, выполняющих специальные функции – антитела иммунной системы, гормоны и др.

Клеточные процессы генетически управляемы. Синтез ферментов, которые управляют метаболизмом, управляется генами. Молекулы РНК переносят от ДНК необходимую для синтеза ферментов информацию и устанавливают связь между генетическими и метаболическими характеристиками клетки. Такая система регуляции

придает клеточным процессам чрезвычайную устойчивость. В последние годы расшифрованы информационные последовательности геномов многих организмов, а также установлена пространственная структура огромного количества различных белков.

Важнейшим свойством клетки как целостной системы является наличие границы – клеточной мембраны. Жидкость внутри клеточной мембраны называется цитоплазмой. Растительные клетки имеют дополнительную жесткую клеточную стенку. Мембрана представляет собой липидный бислой со встроенными молекулами белков, она постоянно находится в активном состоянии – открывается и закрывается, впуская одни вещества внутрь и выпуская другие наружу. Вся деятельность мембраны направлена на поддержание клетки как отдельной сущности и защиту от вредных внешних воздействий.

Все ядерные клетки и большинство бактерий имеют внутренние мембраны, ограничивающие внутреннюю среду клеточных органелл – ядра, митохондрий, хлоропластов в фотосинтезирующих клетках и других.

В цитоплазме клетки, ограниченной мембраной, происходят сложные биохимические процессы, в ходе которых посредством химических реакций и энергетических потоков клетка непрерывно питает, ремонтирует и воспроизводит самое себя. Этот процесс называется метаболизмом. Продукты одних реакций являются субстратами для других, образуя цепи метаболических реакций. Цепи имеют точки разветвления – реакции, продукты которых могут быть использованы не в одной, а в нескольких метаболических цепях. Метаболические процессы представляют собой сети с циклами.

По-видимому, динамика самовоспроизводства и составляет ключевую характеристику живого. В соответствии с этим определением вирусы не являются живыми, так как не обладают собственным метаболизмом. Для производства новой вирусной частицы согласно инструкции, закодированной в ее ДНК, нужно добавить метаболизм живой клетки-хозяина.

Даже у простейших метаболическая сеть чрезвычайно сложна. Большинство процессов ускоряются ферментами и происходят с использованием энергии высокоэнергетических фосфорсодержащих молекул АТФ и других высокоэнергетических соединений. При помощи транспортных РНК сеть каталитических реакций и энергетическая сеть АТФ связаны с геномом, который сам является сложной сетью, где гены прямо или косвенно регулируют работу друг друга.

В своей деятельности клетка постоянно сталкивается с физическим и химическим давлением со стороны окружающей среды. Понимание особенностей физических и химических процессов, протекающих внутри живой системы и на ее границе с окружающей средой, так же важно для понимания основ жизни, как понимание основных биологических генетических информационных программ. Главной трудностью здесь является необходимость интеграции различных областей биологии, физики, химии и математики для описания существующих и создания новых моделей функционирования клетки. Обобщенная модель живой клетки должна описывать биологические и физические модели протекания различных процессов, суммировать базовые математические модели этих процессов.

Институтом математических проблем биологии РАН создан информационно-вычислительный портал обобщенной математической модели эукариотической клетки (www.mathcell.ru). Этот информационно-вычислительный ресурс призван служить своеобразным интерфейсом между биологами и математиками. Одной из компонент является энциклопедия необходимых для моделирования знаний. Она представлена в виде глоссария, поделенного условно на две части – одна объясняет биологические и биохимические термины, другая – физико-математические и биофизические. Это отвечает задаче привлечения математиков к изучению клетки, популярно объяснив

необходимую биологию и просуммировав математику. Биологи с помощью такого интерфейса смогут корректнее ставить перед математиками задачи по моделированию различных аспектов жизни клетки. Благодаря своей архитектуре, ресурс может широко использоваться для образовательных нужд.

Информационный материал представлен в четырех различных вариантах:

- в виде объемной модели животной клетки с возможностями навигации и входами/выходами на уровень тематических статей;
- в виде двухмерного схематического изображения животной и растительной клеток с входами/выходами на уровень тематических статей (рис. 1);
- в виде иерархического дерева тематических статей;
- в виде авторских обзоров по отдельным темам моделирования клетки, сделанных ведущими специалистами в соответствующей области.

Объемная модель животной клетки реализуется специальным программным обеспечением. Трехмерная модель эукариотической животной клетки призвана служить наглядной иллюстрацией материала, описанного тематическими статьями, предоставляя пользователю возможность виртуального экскурса внутрь живой клетки в реальном масштабе времени и являясь, таким образом, наглядным справочником. Будучи интерактивной, модель может исполнять функцию альтернативного трехмерного интерфейса ко всем материалам, представленным на сайте. Это выводит проект на передний край развития компьютерных технологий, поскольку в настоящее время методы представления интерфейсных элементов в трехмерном пространстве являются одним из приоритетных направлений в разработке интерфейсов компьютерных операционных систем и прикладных программ.

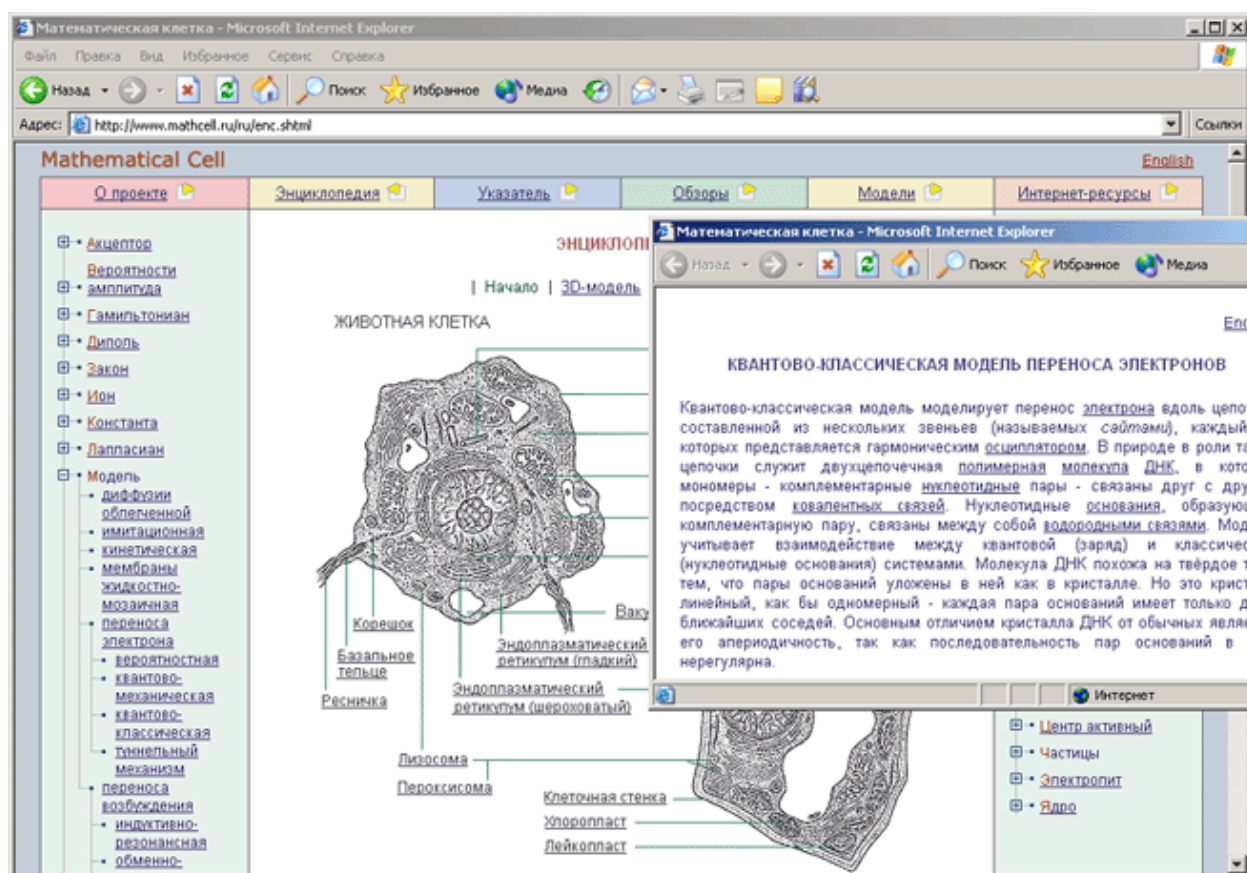


Рис. 1. Иерархическая структура информационной составляющей проекта «Математическая клетка».

В первом варианте реализации данной модели была обеспечена real-time визуализация первого масштабного уровня эукариотической клетки. Пользователь может видеть отдельные клеточные органеллы, свободно «перемещаясь» внутри клеточного пространства (управляя виртуальной камерой, помещенной внутрь модели клетки). Каждая представленная в модели органелла в этом варианте реализации являлась интерфейсным элементом, при воздействии на который осуществлялся переход на соответствующую страницу сайта, описывающую данную органеллу. Эта реализация модели основывалась на программном обеспечении фирмы Anark, что позволяло ей работать как в режиме стационарной программы, загружаемой пользователем на свой персональный компьютер, так и непосредственно в окне Internet Explorer.

Во втором варианте реализации интерактивной трехмерной модели эукариотической клетки была заметно усложнена сама объемная модель. Это позволило значительно увеличить достоверность модели, а рост мощности современной вычислительной техники позволил нивелировать значительно возросший объем обрабатываемых в реальном масштабе времени данных. Построение объемной модели проводилось в программе трехмерного моделирования/анимации MAXON Cinema 4D R9 (рис. 2).

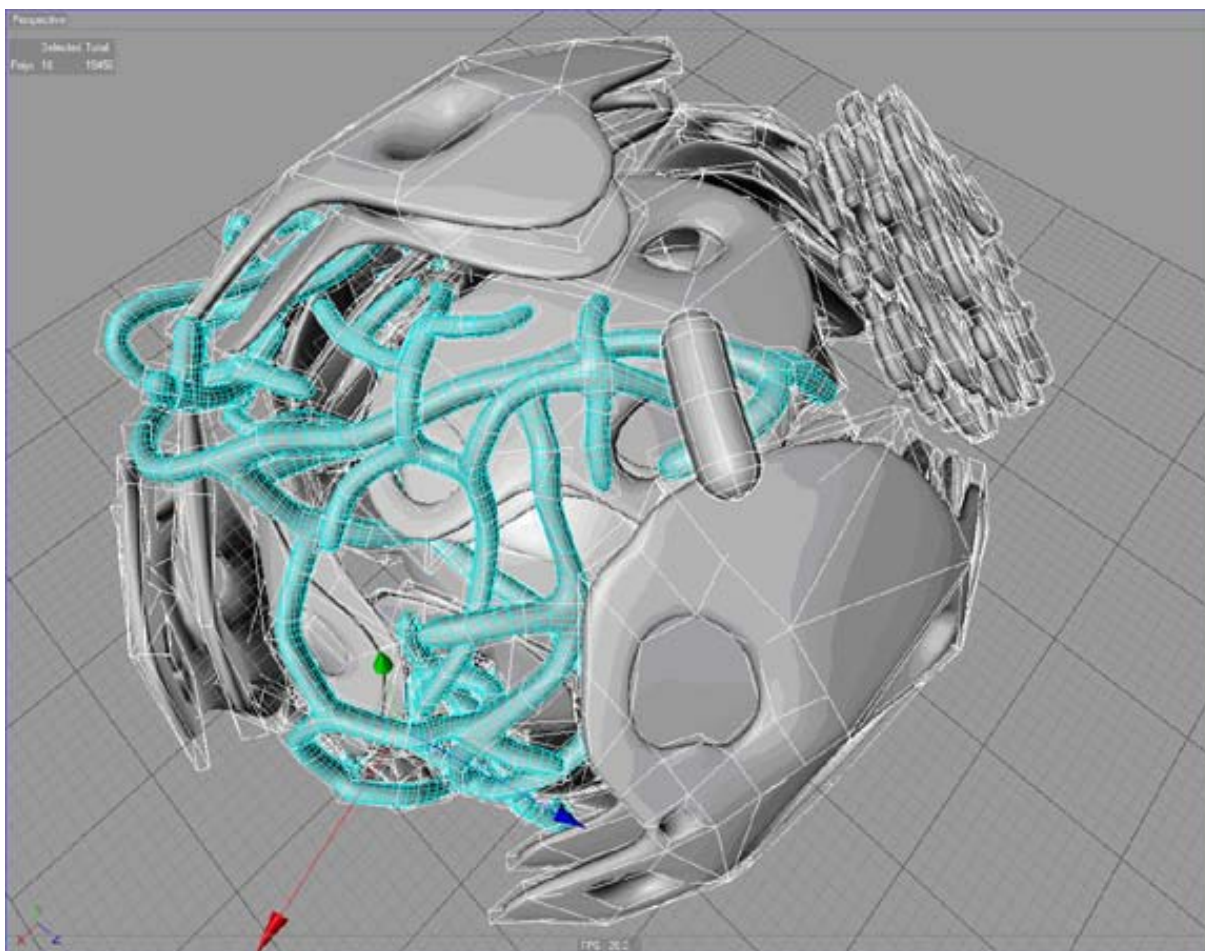


Рис. 2. Трехмерная модель клетки, построенная в среде программы MAXON Cinema 4D R9.

Для визуализации трехмерной модели в реальном масштабе времени и построения логики программы использовался свободно распространяемый пакет 3D-моделирования/анимации – Blender 2.36 (рис. 4). В данной модели было реализовано два ключевых нововведения.

Во-первых, была введена возможность двунаправленного ступенчатого изменения масштаба модели. На примере митохондрии был реализован переход из общеклеточного уровня на уровень отдельной клеточной органеллы. Для этого были сделаны специальные трехмерные модели клеточных органелл со значительно более высокой геометрической детализацией (рис. 3). Это позволит пользователю значительно более подробно изучить внутреннее устройство каждой из органелл.

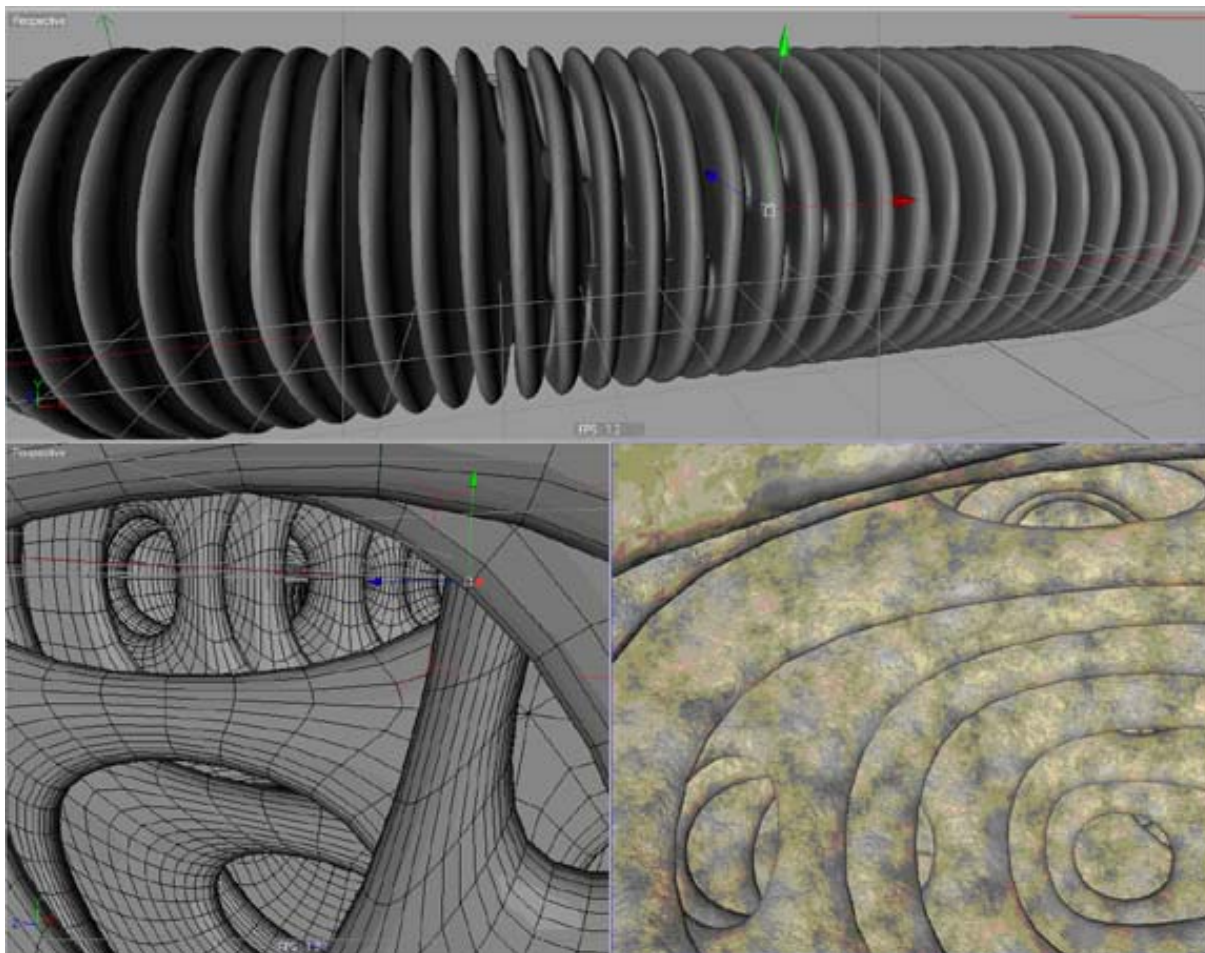


Рис. 3. Детализованная модель клеточной органеллы – митохондрии, построенная в программе MAXON Cinema 4D R9.

Во-вторых, был введен расчет столкновений виртуальной камеры, управляемой пользователем, с трехмерными моделями клеточных органелл. Расчет производится в реальном масштабе времени. Данная возможность является неотъемлемой частью любой имитационной модели, поскольку наилучшим образом позволяет пользователю ощутить основополагающие законы физики, соблюдающиеся и в рамках данной виртуальной среды и таким образом значительно усилить эффект присутствия наблюдателя в смоделированном виртуальном пространстве.

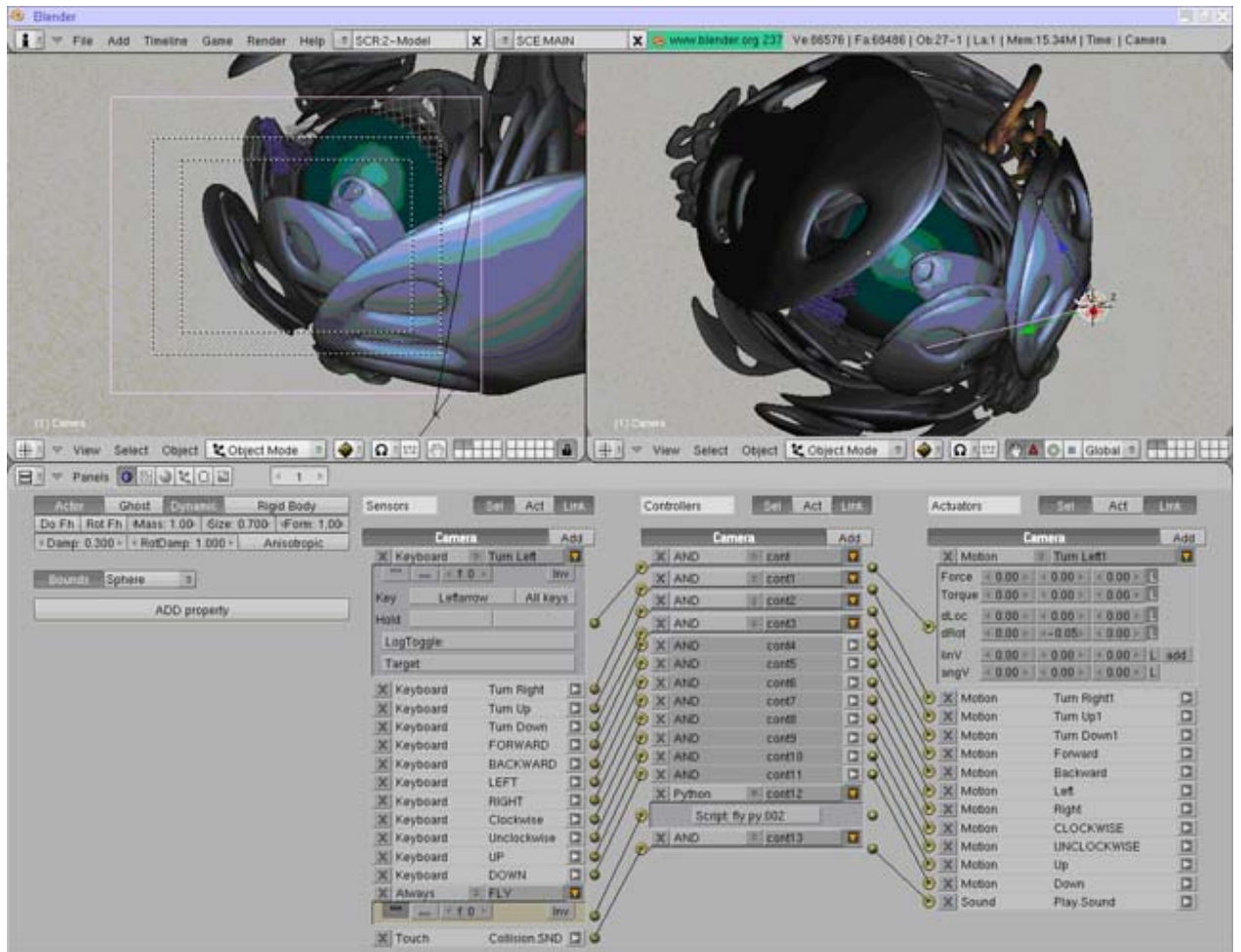


Рис. 4. Управляющая логика интерактивной трехмерной модели эукариотической клетки, созданная в программе Blender v.3.36.

Помимо вышесказанного, в модель был введен ряд усовершенствований, направленных на улучшение ее функциональности. В частности, были значительно расширены возможности навигации в трехмерном пространстве (рис. 5).

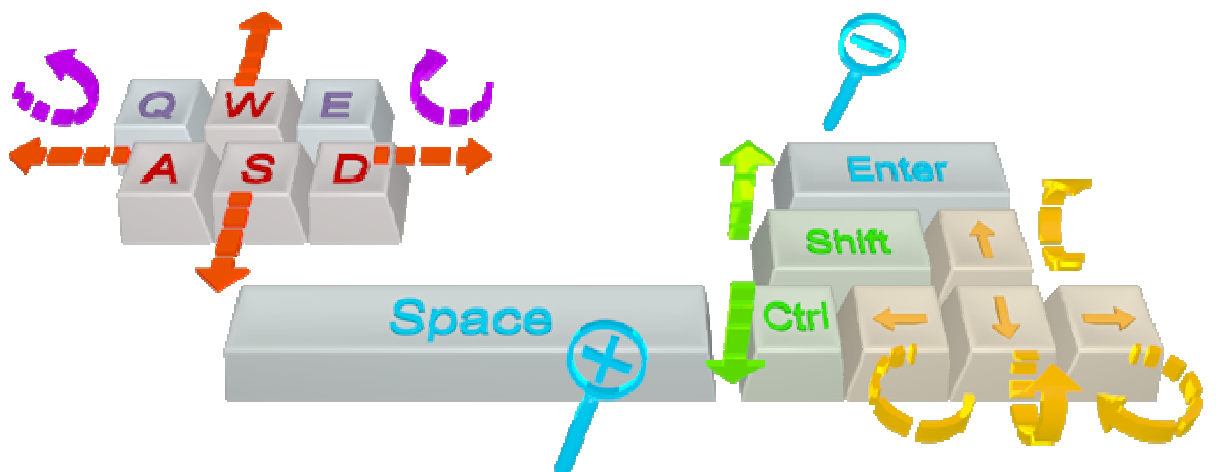


Рис. 5. Иллюстрация способов навигации по трехмерной модели клетки.

Некоторые элементы модели были текстурированы. Были добавлены звуковые эффекты, усиливающие эффект обратной связи с пользователем. Из-за особенностей использованного программного обеспечения данная модель предлагается лишь в виде программы, загружаемой на локальный компьютер пользователя.

К настоящему времени возможности используемого ранее программного обеспечения были исчерпаны. На фоне значительного накопления исходного материала (трехмерные модели органелл, фрагменты программного кода, текстуры и т.п.) был осуществлен переход на использование коммерческой системы визуального программирования трехмерных интерактивных программ Quest3D v.3.6 фирмы Act-3D. В рамках этого перехода геометрия трехмерной модели была значительно оптимизирована. Число использованных полигонов в модели клетки было доведено (уменьшено) до 115000 с одновременным увеличением геометрической детализации и улучшением визуального качества модели.

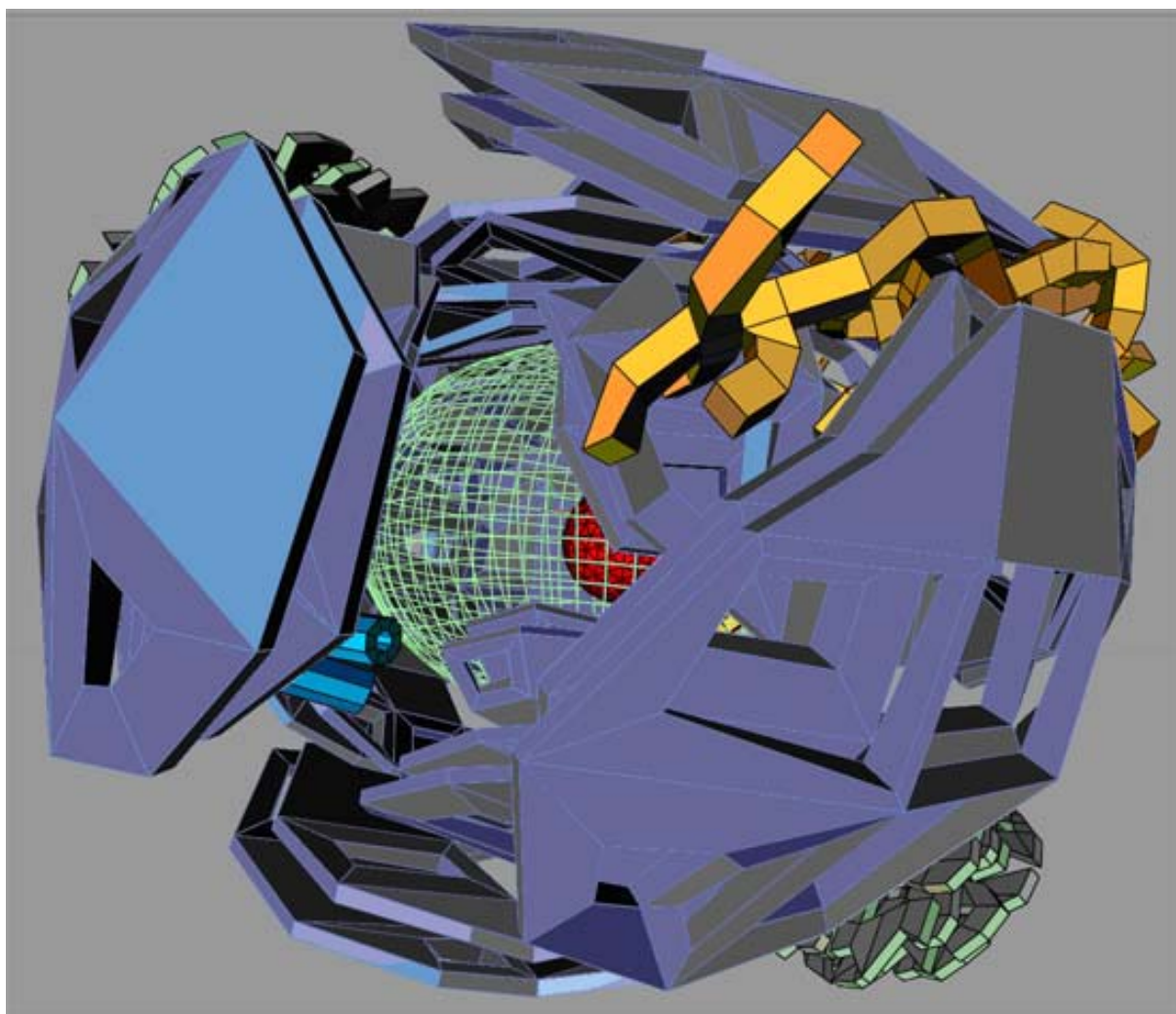


Рис. 6. Упрощенная геометрическая модель клетки, используемая для расчета столкновений.

Поскольку точный расчет столкновений виртуального наблюдателя, находящегося в окружении моделей клеточных органелл, с объектами в реальном масштабе времени представляет собой довольно емкую с точки зрения вычислительных ресурсов задачу, то для существенного ускорения этих расчетов была создана специальная геометрическая модель клетки с экстремально низкой детализацией. Такая модель существует параллельно основной и не визуализируется, а лишь участвует в расчете

столкновений управляемой пользователем виртуальной камеры, а так же некоторых небольших подвижных объектов (например, пузырьков аппарата Гольджи). Число полигонов в этой модели меньше почти в 10 раз и составляет 13500. Программная логика, отвечающая за базовую визуализацию моделей, навигацию и систему расчета столкновений описана с помощью блок-схемы в среде программного пакета Quest3D (рис. 7).

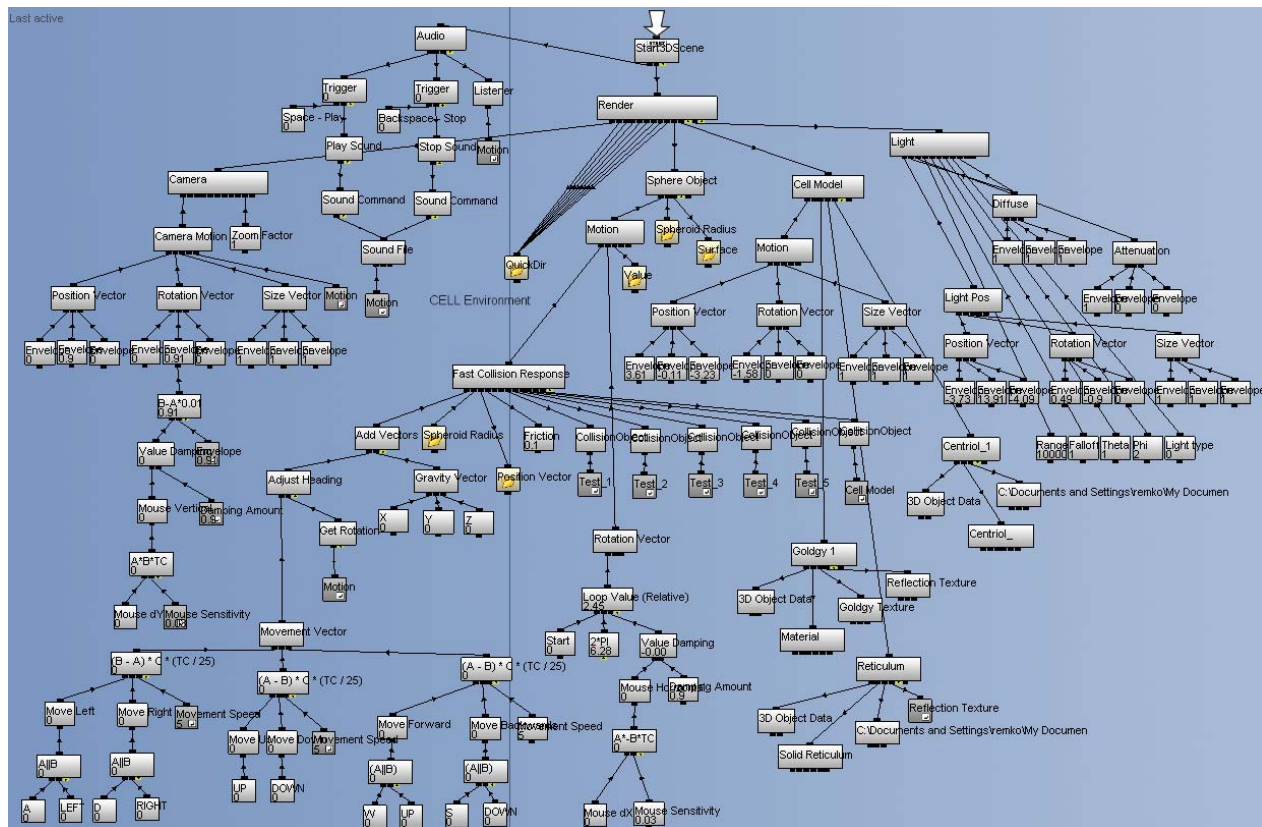


Рис. 7. Блок-схема, созданная в среде Quest3D, определяющая функциональность программы интерактивной трехмерной модели эукариотической клетки.

Использование Quest3D позволит снять практически все ранее существовавшие ограничения и в полной мере использовать все новейшие технологии трехмерной визуализации в реальном времени. Результирующая программа может быть представлена как в виде самостоятельной компьютерной программы, загружаемой пользователем на его локальный персональный компьютер, так и в виде модуля, визуализируемого непосредственно в окне одного из наиболее распространенных интернет-браузеров (Internet Explorer, Firefox и др.).

Для расширения функциональности собственно самой модели клетки, а также состава программно-аппаратных платформ, способных служить средой для исполнения трехмерной интерактивной модели живой клетки, все созданные нами графические и алгоритмические ресурсы были скомпилированы в новую модель эукариотической клетки с помощью коммерческого программного комплекса Virtools v.4.0. Полученная программа функционально аналогична программе, полученной в среде Quest3D.

В настоящее время выполнена подготовка текстурных карт и программирование шейдеров, предназначенных для имитации некоторых ключевых свойств поверхностей таких органелл, как ядро, шероховатый эндоплазматический ретикулум и т.п. По мере подготовки моделей и текстур в ближайшее время планируется:

- увеличение числа шагов масштабирования;
- информационное наполнение трехмерной модели клетки;

- введение элементов анимации, позволяющих демонстрировать процессы, происходящие в клетке;
- программирование пользовательского интерфейса, позволяющего взаимодействовать не только с объектами модели, но и со всеми материалами сайта;
- программирование интерфейса взаимодействия пользователя модели с европейской грид-инфраструктурой.

Тематические статьи снабжены рисунками, схемами, перекрестными ссылками на другие статьи. Кроме того, в них представлены ссылки на интернет-ресурсы и литературу, где можно получить дополнительную информацию по теме статьи. Язык представления материала максимально упрощен, глубина подачи материала соответствует общеобразовательному уровню знаний в области биологии и химии. На уровень тематических статей есть выходы с любого уровня информационно-программной среды – в обзорах, в описаниях моделей, даже в программно-пользовательских интерфейсах есть активные ссылки на соответствующие тематические статьи. В текущей версии портала выставлено около 500 тематических статей на русском и английском языках.

Информационно-вычислительная среда интегрирует различные информационные и программные ресурсы. Среди них специализированные базы данных, аккумулирующие разного рода информацию о компонентах клетки, ее функционировании и программных разработках в математическом моделировании процессов клетки. Сделаны активные ссылки на мировые информационные ресурсы. Кроме того, собрана и выложена информация (активные ссылки) о научных интернет-изданиях (журналах), публикующих материалы по клеточной биологии. В представленной на сегодняшний день версии информационной части портала более подробно описаны математические модели биохимических и биофизических процессов, происходящих в митохондриях и мембранах. В частности, описаны классическая, вероятностная, квантово-механическая и квантово-классическая модели транспорта электронов в дыхательной цепи, расположенной на внутренней мембране митохондрий и хлоропластов (у растений). Математическое моделирование переноса электронов в комплексах молекул-переносчиков, локализованных в энергосопряженных мембранах митохондрий и фотосинтезирующих органелл – хлоропластов растений и хроматофоров фотосинтезирующих бактерий – имеет большое значение для изучения катаболизма клетки в общем и процесса окислительного фосфорилирования в частности.

Коллекция моделей MathCell представляет собой депозитарий наших собственных моделей. В настоящее время она включает в себя:

- квантово-классическую модель переноса электрона вдоль молекулы биополимера;
- модель переноса электрона в фотосинтетической мембране;
- расчет энергии растворения биомолекул в воде методом Монте-Карло.

Модель переноса заряда вдоль фрагмента ДНК (рис.8). Модель позволяет исследовать перенос заряда (электрона либо дырки) в нуклеотидных последовательностях различной длины и состава при предписанной температуре окружающего раствора на основе комбинированного квантово-классического подхода. Фрагмент ДНК из N пар оснований рассматривается как цепочка из N сайтов; изменение параметров разных сайтов позволяет моделировать заданную последовательность нуклеотидов. Движение заряда описывается уравнением Шредингера, а колебания сайтов (нуклеотидных пар) – классическими уравнениями движения. Для учета температуры окружающей среды в классические уравнения движения сайтов добавлен ланжевеновский член (случайная сила с определенными характеристиками), моделирующий температурные флуктуации окружающей среды. При моделировании динамики переноса заряда в ДНК проводятся расчеты множества

отдельных реализаций – траекторий из различных начальных данных для классической подсистемы и с разными наборами значений для случайных сил, с последующим осреднением по реализациям.

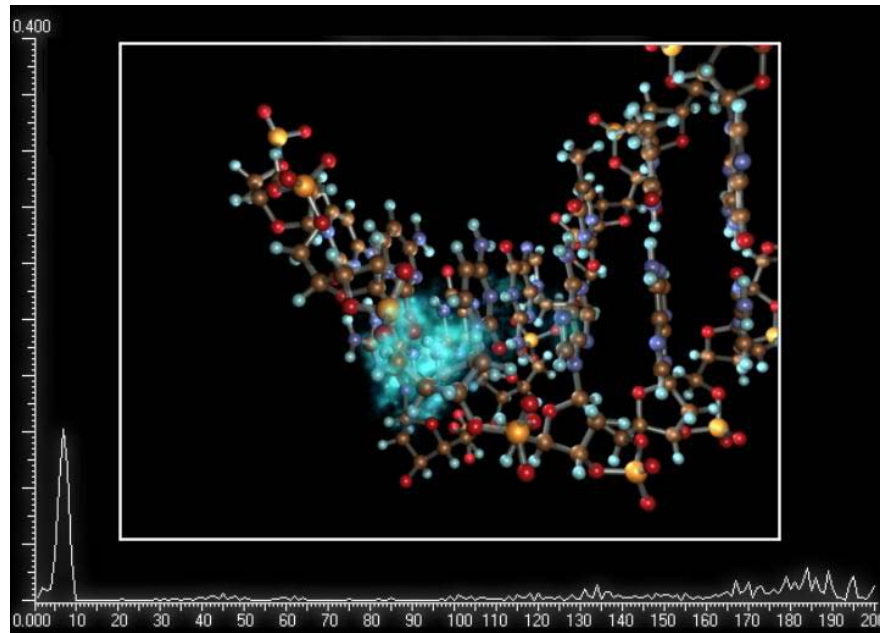


Рис. 8. График распределения вероятностей нахождения дырки по сайтам однородной G/C-последовательности с донором и акцептором в некоторый момент модельного расчета. Расчеты показывают, что заряд долгое время локализован на сайтах донора; динамика переноса зависит от параметров донора и акцептора.

Численные эксперименты по переносу заряда вдоль заданного фрагмента ДНК позволяют оценить время переноса, найти место конечной локализации заряда, рассчитать подвижность заряда в такой последовательности.

Модель переноса электрона в фотосинтетической мембране (рис. 9). Модельная сцена представляет собой область в трехмерном пространстве, включающую несколько компартментов: участок двойной фотосинтетической мембраны, межмембранное пространство (люмен) и область снаружи мембраны (строму). В модель включены мобильные переносчики электрона (пластоцианин в люмене, ферредоксин в строме и пластохинон в мембране), а также встроенные в мембрану белковые комплексы фотосистем I и II и цитохромный *b6/f* комплекс.

Параметрами модели являются геометрические размеры моделируемого участка фотосинтетической мембраны, и количество (плотность) комплексов. Движение мобильных переносчиков описывается броуновской динамикой. Задаются коэффициенты диффузии переносчиков. В модели учитываются столкновения между переносчиками и белковыми комплексами. Реакции между переносчиками и комплексами описываются через радиус взаимодействия d и вероятность реакции p . Если переносчик электрона подходит к сайту посадки на комплексе на расстояние, меньшее d , с вероятностью p возможна стыковка и реакция передачи электрона.

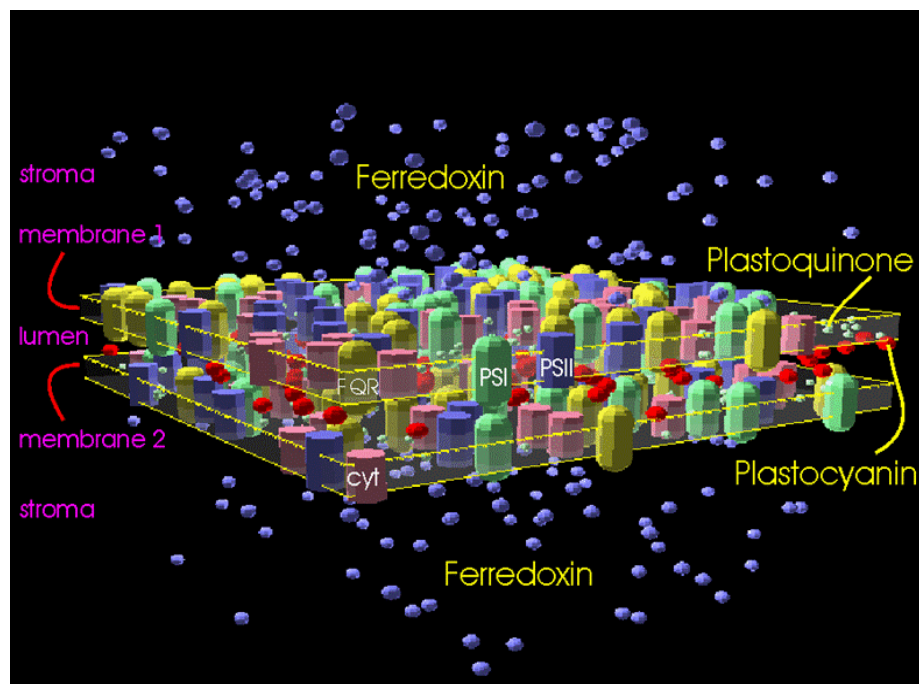


Рис. 9. Схематическое изображение модельной сцены для имитации переноса заряда в фотосинтетической мембране. PSI и PSII соответственно обозначают комплексы фотосистем I и II, cyt - цитохромные комплексы.

Расчет энергии растворения биомолекул в воде методом Монте-Карло. Программный комплекс предназначен для изучения энергетических аспектов взаимодействия небольших (несколько десятков атомов) молекул биологически активных веществ или их фрагментов с водным окружением.

Компьютерное моделирование методом Монте-Карло позволяет рассчитывать среднюю энергию исследуемой молекулярной системы по статистически значимой выборке её конфигураций при заданной температуре. Генерация (выборка) молекулярной конфигурации выполняется посредством процедуры Метрополиса [30], осуществляющей стохастический процесс (марковская цепь) блуждания в координатном пространстве изучаемой модели [30–32]. Энергия межмолекулярных взаимодействий рассчитывается с помощью атом-атомных потенциальных функций [31, 32].

Моделируемая система представляет собой ячейку в форме прямоугольного параллелепипеда, заполненную молекулами воды и содержащую исследуемую молекулу в центральной части модели. Размеры ячейки и количество молекул воды выбираются (автоматически) таким образом, чтобы энергия взаимодействия растворенной молекулы с периферийными водными молекулами была существенно меньше энергии водородных связей. Для устранения краевых эффектов на ячейку налагаются периодические граничные условия.

Энергия растворения (для бесконечного разбавления) будет вычисляться как разность средней энергии моделируемой системы (растворенная молекула + N молекул воды) и средней энергии водной системы без исследуемой молекулы (в несколько меньшей ячейке те же N молекул воды).

Для ускорения счета и получения более надежных результатов рекомендуется разбивать большие молекулы на небольшие и, по возможности, электронейтральные фрагменты. Расчет для большой молекулы может затянуться на десятки часов, поэтому в данной версии программ имеются количественные ограничения на число атомов (максимальное число фрагментов 20, максимальное число атомов во фрагменте 30, не более 2000 молекул воды).

Данное программное обеспечение может использоваться в различных целях - фундаментальных исследованиях и прикладных разработках лекарственных средств. Например, с его помощью можно получать относительные данные для оценки энергии растворения молекул из гомологичных рядов, а по температурным зависимостям оценивать энтальпийные и энтропийные эффекты растворения.

Для каждой модели есть демо-версия программного обеспечения, кроме того, пользователю предоставляется возможность расчетов в ГРИД-инфраструктуре, что дает возможность учитывать в моделях больше деталей, т.е. усложнять модели. Наши модели используют различные математические аппараты - решение дифференциальных уравнений, имитационное моделирование, моделирование методом Монте-Карло. Для организации таких расчетов создано специальное программное обеспечение, состоящее из двух частей - интерфейс пользователя и система организации вычислений в ГРИД.

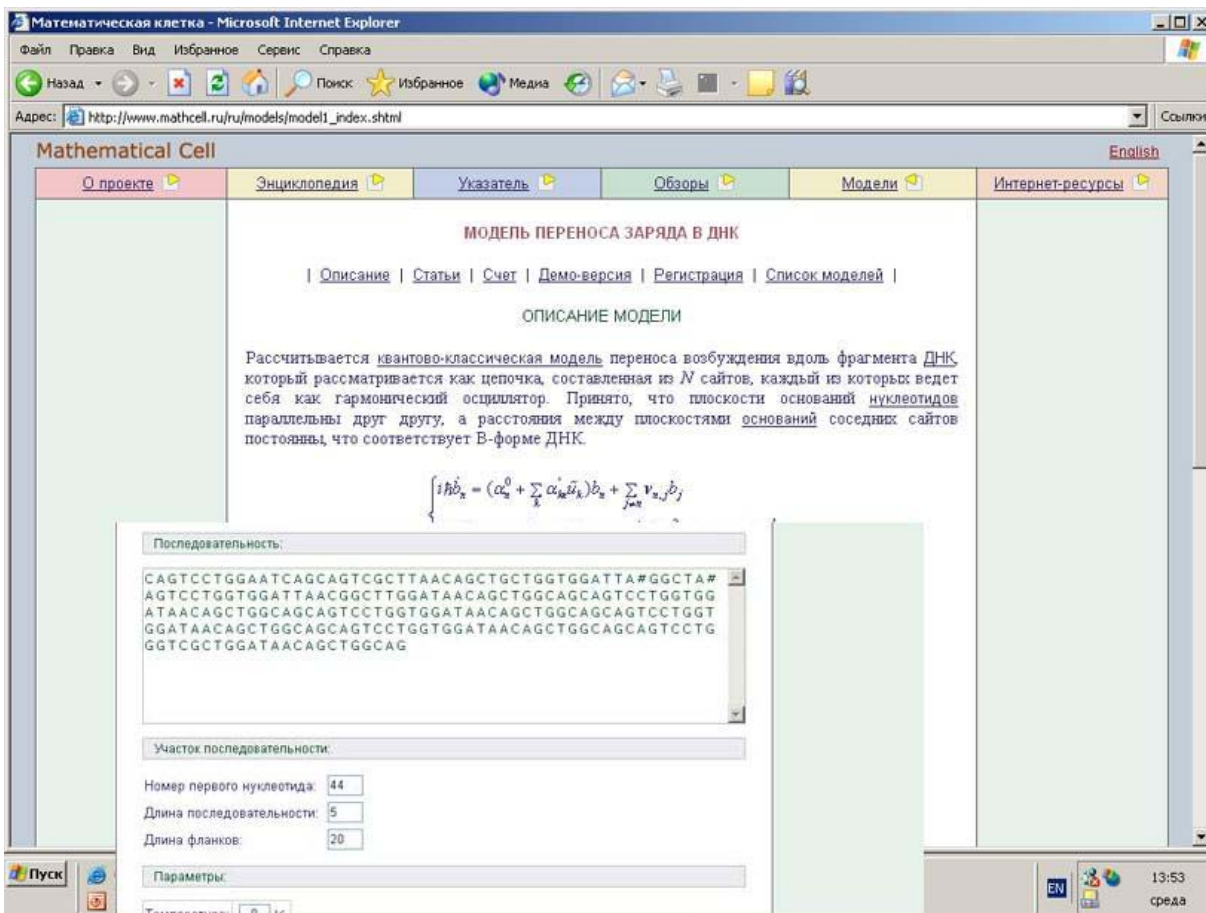


Рис.10. Интерфейс, созданный для работы с моделью переноса электронов в ДНК.

Интерфейс пользователя (рис.10) реализует интерактивный процесс выбора модели и позволяет задавать параметры моделей из множества допустимых значений. Это программное обеспечение также визуализирует выходные данные расчетов и представляет их в удобной для восприятия пользователем форме. Модели коллекции MathCell снабжены подробнейшими описаниями на двух языках – русском и английском, с объяснением всех параметров моделей и их допустимыми значениями.

Особых усилий потребовало создание системы запуска заданий в ГРИД-инфраструктуру. В настоящее время на нашем портале можно запустить задание на счет, а через некоторое время получить результаты счета. Система запуска заданий работает с внутренней базой данных заданий, ставит их в свою внутреннюю очередь, задания из очереди запускаются в ГРИД. Эта система универсальна, она работает с

любой задачей, что делает размещение новых задач в ГРИД тривиальной процедурой. Если запуск задания в ГРИД-инфраструктуру не получается из-за отсутствия необходимых ресурсов, программа-робот через определенные промежутки времени снова повторяет попытку запуска задачи, она же отслеживает статус задачи и осуществляет получение результатов по окончании счета. Пользователь, знакомый с ОС Windows, не обязан знать ничего о структуре и программном обеспечении ГРИД, командах ОС Linux.

В планах коллектива интеграция в коллекцию MathCell других расчетных моделей, которые традиционно используют в своих расчетах суперкомпьютерные мощности кластера ИМПБ РАН. Такая интеграция подразумевает создание удобного веб-интерфейса для задания параметров моделирования, запуска на счет, представления и интерпретации результатов расчетов.

Работа выполнена при частичной поддержке РФФИ, гранты №№ 07-07-00313, 06-01-08039, 06-07-89274, 07-07-00280, 06-07-89303, 07-01-00490, а также средств международного проекта EGEE II 6-й Рамочной программы Европейской комиссии (контракт №031688).

ЛИТЕРАТУРА

1. С. Гринченко, С. Загускин. *Механизмы живой клетки: алгоритмическая модель*. М.: Наука. 1989. 232 с.
2. D. Normile. Building Working Cells 'In Silico'. *Science*. 1999. **284**. 80-81.
3. R. Triendl. Computerized role models. *Nature*. 2002. **417**. 7
4. E-CELL Project Home Page, <http://www.e-cell.org/ecell/>
5. Tomita M, Hashimoto K, Takahashi K, Shimizu TS, Matsuzaki Y, Miyoshi F, Saito K, Tanida S, Yugi K, Venter JC, Hutchison CA 3rd. E-CELL: software environment for whole-cell simulation. *Bioinformatics*. 1999. **15**(1). 72-84.
6. National Resource for Cell Analysis and Modeling, <http://vcell.org/>
7. BioUML - open source Java framework for systems biology, <http://www.biouml.org/>
8. Center for Cell and Virus Theory. Cell Modelling Page, <http://biodynamics.indiana.edu/CellModeling>
9. Ortoleva P, Berry E, Brun Y, Fan J, Fontus M, Hubbard K, Jaqaman K, Jarymowycz L, Navid A, Sayyed-Ahmad A, Shreif Z, Stanley F, Tuncay K, Weitzke E, Wu LC. The Karyote physico-chemical genomic, proteomic, metabolic cell modeling system. *OMICS*. 2003. **7**(3). 269-283.
10. BioNetGen – Biological Network Generator, <http://cellsignaling.lanl.gov/bionetgen>
11. J. R. Faeder, M. L. Blinov, B. Goldstein and W. S. Hlavacek (2005) Rule-based modeling of biochemical networks. *Complexity* **10**. 22-41.
12. Blinov ML, Faeder JR, Goldstein B, Hlavacek WS. BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics*. 2004. **20**(17). 3289-3291.
13. Sauro HM, Hucka M, Finney A, Wellock C, Bolouri H, Doyle J and Kitano H. Next generation simulation tools: the Systems Biology Workbench and BioSPICE integration. *OMICS*. 2003. **7**(4). 355-372.
14. P. K. Dhar, T. Ch. Meng, S. Somani, Li Ye, K. Sakharkar, A. Krishnan, A. B. M. Ridwan, S. Ho Kok Wah, M. Chitre, Zhu Hao. Grid Cellware: the first grid-enabled tool for modelling and simulating cellular processes. *Bioinformatics*. 2005. **21**(7). 1284–1287

15. Cellware: A New Modeling and Simulation Tool for Modeling Cellular Transactions, <http://www.bii.a-star.edu.sg/research/sbg/cellware/>
16. Systems Biology Markup Language, <http://sbml.org/index.psp>
17. Extensible Markup Language (XML) 1.0 (Fourth Edition), <http://www.w3.org/TR/REC-xml/>
18. World Wide Web Consortium, <http://www.w3.org/>
19. A. Finney and M. Hucka. Systems Biology Markup Language: Level 2 and Beyond. *Biochem. Soc. Trans.* 2003, **31**. 1472-1473.
20. A. Finney. Developing SBML Beyond Level 2: Proposals for Development. *Lecture Notes in Computer Science*. 2005. **3082**. 242-247.
21. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, Dronov S, Gilles ED, Ginkel M, Gor V, Goryanin II, Hedley WJ, Hodgman TC, Hofmeyr JH, Hunter PJ, Juty NS, Kasberger JL, Kremling A, Kummer U, Le Novère N, Loew LM, Lucio D, Mendes P, Minch E, Mjolsness ED, Nakayama Y, Nelson MR, Nielsen PF, Sakurada T, Schaff JC, Shapiro BE, Shimizu TS, Spence HD, Stelling J, Takahashi K, Tomita M, Wagner J, Wang J; SBML Forum. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*. 2003. **19**(4). 524-531.
22. Shapiro BE, Hucka M, Finney A, Doyle J. MathSBML: a package for manipulating SBML-based biological models. *Bioinformatics*. 2004. **20**(16). 2829-2831.
23. CellML Home Page, <http://www.cellml.org/>
24. Cuellar, A.A., Lloyd, C.M., Nielsen, P.F., Bullivant, D.P., Nickerson, D.P. and Hunter, P.J. An Overview of CellML 1.1, a Biological Model Description Language. *SIMULATION: Transactions of The Society for Modeling and Simulation International*. 2003. **79**(12). 740-747.
25. Lloyd, C.M., Halstead, M.D.B. and Nielsen, P.F. CellML: its future, present and past. *Progress in Biophysics and Molecular Biology*. 2004. **85**(2-3). 433-450.
26. Schilstra MJ, Li L, Matthews J, Finney A, Hucka M, Le Novère N. CellML2SBML: conversion of CellML into SBML. *Bioinformatics*. 2006. **22**(8). 1018-1020.
27. The BioModels Database, <http://www.ebi.ac.uk/biomodels/>
28. Le Novere N, Bornstein B, Broicher A, Courtot M, Donizelli M, Dharuri H, Li L, Sauro H, Schilstra M, Shapiro B, Snoep JL, Hucka M. BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res.* 2006. **34**. D689-D691
29. BioPAX : Biological Pathways Exchange. <http://www.biopax.org/>
30. Фишер И.З. *Статистическая теория жидкостей*. М.: ГИФМЛ. 1961. 280 с.
31. Poltev V.I., Grokhlina T.I. and Malenkov G.G. *J. Biomol. Struct. Dyn.* 1984. **2**. 413-429.
32. Poltev V.I., Malenkov G.G., Gonzalez E.J., Teplukhin A.V., Rein R., Shibata, M. and Miller J.H. *J. Biomol. Struct. Dyn.* 1996. **13**. 717-725.

Материал поступил в редакцию 29.11.2007, опубликован 29.12.2007.